한국음향학회지 제21권 제4호 pp. 387~393 (2002)

혼합 은닉필터모델 (HFM)을 이용한 비정상 잡음에 오염된 음성신호의 향상

Speech Enhancement Based on Mixture Hidden Filter Model (HFM) Under Nonstationary Noise

강 상 기*, 백 성 준**, 이 기 용***, 성 굉 모****
(SangKi Kang*, SeongJoon Baek**, Ki Yong Lee***, Koeng-Mo Sung****)

*서울대학교 대학원 전기·컴퓨터공학부, **전남대학교 전자컴퓨터정보통신공학부
***숭실대학교 정보통신과, ****서울대학교 전기·컴퓨터공학부

(접수일자: 2002년 1월 14일; 채택일자: 2002년 4월 8일)

비정상 잡음에 오염된 음성신호의 향상을 위하여 혼합 은닉필터모델 (HFM: Hidden Filter Model)에 기초한 기법을 제안하였다. 오염된 음성신호를 선형상태방정식으로 모델링하고 파라미터는 마코프 모델에 따른다고 가정하였다. 이 파라미터들은 잡음에 오염되지 않은 학습신호로부터 추정할 수 있다. 추정과정은 혼합 상호복합 모델 (IMM: Interacting Multiple Model)에 기초하여 이루어지며, 음성신호의 추정값은 상호작용하는 병렬의 칼만 필터들의 가중합으로 주어진다. 실험결과로부터 제안한 방법의 성능이 기존의 방법에 비해 개선되었음을 확인할 수 있었다.

핵심용어: 혼합 은닉필터모델, 혼합 상호복합모델, 다중 칼만필터, 비정상 잡음
투고분야: 음성처리 분야 (2,3)

The enhancement technique of noisy signal using mixture HFM (Hidden Filter Model) are proposed. Given the parameters of the clean signal and noise, noisy signal is modeled by a linear state-space model with Markov switching parameters. Estimation of state vector is required for estimating original signal. The estimation procedure is based on mixture interacting multiple model (MIMM) and the estimator of speech is given by the weighted sum of parallel kalman filters operating interactively. Simulation results showed that the proposed method offers performance gains relative to the previous results with slightly increased complexity.

Keywords: *Miture HFM (hidden filter model), Mixture IMM (interacting multiple model), Multiple kalman filter, Nonstationary noise*
ASK subject classification: *Speech signal processing (2,3)*

I. 서 론

음성향상은 잡음에 의해 오염된 음성신호를 복구하는 것 자체로도 음성의 인지도를 향상시킨다는 점에서 의미가 있지만, 잡음환경에서의 음성인식이나 음성 부호화를 위한 전처리 과정으로서도 중요한 의미를 가지고 있다. 기존의 음성향상을 위한 방법들은 크게 스펙트럼 차감법 [1], 자기회귀 (AR: autoregressive) 모델링과 평균 최대 (EM: Expectation Maximization)알고리즘에 기초한 방법[2]과 은닉마코프모델 (HMM: Hidden Markov Model)에 기초한 방법[3-5] 등으로 나눌 수 있다.

그 중 HMM에 기초한 대부분의 방법들은 음성신호를 일정한 길이의 프레임으로 나누어서 처리하였다. 따라서

책임저자: 강상기 (sangkikang@samsung.com)
151-742 서울시 관악구 신림동 산56-1번지
서울대학교 전기컴퓨터공학부
(전화: 02-880-7263; 팩스: 02-882-4657)

음소가 변하는 전이구간들을 잘 모델링할 수 없다는 문제가 있었다[6]. 그리고 이들 방법은 서로 다른 프레임간의 음성신호가 통계적으로 서로 독립적인 신호원에 의해 생성된다는 가정을 전제로 했으나 실제로 음성신호의 경우 인접한 프레임간에는 매우 큰 상관관계가 존재한다[3]. 또한 이들은 대부분이 순환적인 알고리즘이 아니므로 실시간 처리에 적합치 않고 저 지연 음성부호화기의 전처리 과정으로 사용될 경우 프레임에 의한 지연이 추가적으로 발생하게 된다. 최근에 이러한 기존의 HMM에 기초한 방법들의 문제점을 해결하기 위하여 HFM을 이용하여 신호를 향상시키는 방법이 제안되었다[7]. 여기서 음성신호는 마코프 모델에 따라서 변하는 파라미터를 갖는 시변 AR 프로세서로 모델링된다. 그러나 기존방법에서는 마코프 모델의 각 상태에 하나의 가우시안 AR 모델을 사용하는 단일 HFM이 사용되는데 비해[8] 본 논문에서는 신호의 좀더 정확한 모델링을 위해 각 상태에 여러 개의 가우시안 AR 모델을 사용하는 혼합 HFM을 사용하였다. 학습 음성신호로부터 추정한 혼합 HFM의 파라미터와 묵음구간으로부터 얻은 HFM의 파라미터를 추정하고, 혼합 IMM알고리즘을 부가적인 비정상 잡음에 오염된 음성신호의 향상에 적용한다.

II. 음성신호의 모델링

잡음없는 음성신호는 마코프 모델에 따라서 변하는 파라미터를 갖는 AR 모델인 혼합 HFM에 의해서 모델링된다. L개의 상태에 해당하는 모델 $s(t)\in\{1,\cdots,L\}$, M개의 혼합 (mixture)에 해당하는 $m(t)\in\{1,\cdots,M\}$와 상태천이 행렬 $a_{s(t-1)|s(t)}$, 혼합 확률 $c_{m(t)|s(t)}$를 갖는 일차의 마코프 모델을 생각하자. 시간 t에서의 상태 $s(t)$에서 혼합 $m(t)$를 갖는다면 잡음없는 음성신호는 다음과 같이 표현된다.

$$y(t) = \boldsymbol{B}^T_{m(t)|s(t)}\, \boldsymbol{y}(t-1) + e_{m(t)|s(t)}(t) \tag{1}$$

여기서 $\boldsymbol{B}_{m(t)|s(t)}=[b_{m(t)|s(t)}(1)\cdots b_{m(t)|s(t)}(y)]^T$는 상태 $s(t)$에서 혼합 $m(t)$일 때의 AR 계수 벡터이고 $\boldsymbol{y}(t-1)=[y(t-1)\cdots y(t-p)]^T$는 p개의 데이터로 구성된 신호의 벡터이며, $e_{m(t)|s(t)}(t)$는 평균이 0이고 분산이 $\sigma^2_{m(t)|s(t)}$인 가우시안 프로세스이다. HFM의 파라미터 $\lambda_y=\{\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{B}, \sigma^2\}$은 잡음에 오염되지 않은 학습음성신호 $y_1^T=\{y(1),\cdots,y(T)\}$로부터 Baum-Welch 알고리즘을 사용하여 추정할 수 있다[8]. 여기서 $\boldsymbol{a}=\{a_{ij}\}$, $\boldsymbol{c}=\{c_{k|j}\}$, $\boldsymbol{B}=\{\boldsymbol{B}_{k|j}\}$, $\sigma^2=\{\sigma^2_{kj}\}$이고 $i, j=1,\ldots,L$ 그리고 $k=1,\ldots,M$이다. 모델의 초기값 λ_y에서 시작하여 다음의 보조함수 $Q(\lambda'_y, \lambda_y)$를 최대화하는 λ'_y를 구하고 다시 λ'_y를 λ_y로 놓고 위의 과정을 반복하게 되면 HFM 파라미터를 구할 수 있게 된다.

$$\begin{aligned}
Q(\lambda'_y\ \lambda'_y) &= \sum_{t=1}^{T}\sum_{m(t)}\sum_{s(t)} p_{\lambda_{t-1}}(m(t), s(t)|y(t)) \\
&\quad \times \ln p_{\lambda'_t}(y(t)|m(t), s(t)) \\
&= \sum_{t=1}^{T}\sum_{m(t)}\sum_{s(t)} p_{\lambda_{t-1}}(m(t), s(t)|y(t)) \\
&\quad \times \ln[p(s(t))p(m(t)\mid s(t))p_{\lambda'_t}(y(t)|m(t), s(t)) \\
&= \sum_{t=1}^{T}\sum_{m(t)}\sum_{s(t)} p_{\lambda_{t-1}}(m(t), s(t)|y(t)) \\
&\quad \times [\ln a_{s(t-1)s(t)} + \ln c_{m(t)|s(t)} + \ln D_{m(t)|s(t)}
\end{aligned} \tag{2}$$

여기서

$$\ln D_{m(t)|s(t)} = \frac{1}{\sqrt{2\pi}\sigma_{m(t)|s(t)}}$$

$$\exp\left[-\frac{1}{2\sigma^2_{m(t)|s(t)}}(y(t)-B^T_{m(t)|s(t)}\ \boldsymbol{y}_{(t-1)})^2\right]$$

이고, $p_{\lambda_{t-1}}(m(t), s(t)|y(t))$는 $t-1$에서 신호 $y(t)$가 주어졌을 때 상태 $s(t)$와 혼합 $m(t)$의 사후확률이다. 식 (2)를 $\boldsymbol{a}=\{a_{ij}\}$, $\boldsymbol{c}=\{c_{k|j}\}$, $\boldsymbol{B}=\{\boldsymbol{B}_{k|j}\}$, $\sigma^2=\{\sigma^2_{kj}\}$ 각각에 대해 미분한 값을 0으로 하는 조건을 구하면 다음 식에 의해 구할 수 있다. 여기서 $i, j=1,\ldots,L$ 그리고 $k=1,\ldots,M$이다.

$$a_{ij} = \frac{\left[\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t-1)=i, s(t)=j\mid y(t))\right]}{\left[\sum_{j=1}^{L}\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t-1)=i, s(t)=j\mid y(t))\right]} \tag{3}$$

$$c_{k|j} = \frac{\left[\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t)=j, m_t=k\mid y(t))\right]}{\left[\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t)=j\mid y(t))\right]} \tag{4}$$

$$\boldsymbol{B}_{k|j} = \frac{\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t)=j, m(t)=k\mid y(t))\times y(t)\ \boldsymbol{y}(t-1)}{\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t-1)=i, s(t)=j\mid y(t))\times y(t-1)\ \boldsymbol{y}^T(t-1)} \tag{5}$$

$$\sigma^2_{k|j} = \frac{\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t)=j, m(t)=k\mid y(t))(y(t)-\ B^T_{k|j}\ y(t-1))^2}{\sum_{t=1}^{T} p_{\lambda_{t-1}}(s(t)=j, m(t)=k\mid y(t))} \tag{6}$$

여기서

$p_\lambda(s(t-1)=i, s(t)=j \mid y(t))$, $p_\lambda(s(t)=j, m(t)=k \mid y(t))$는 "순방향-역방향 알고리즘 (forward-backward algorithm)"에 의해 다음과 같이 효과적으로 계산할 수 있다.

$$p_\lambda(s(t-1)=i, s(t)=j \mid y(t)) = \alpha_{t-1}(i)\, a_{ij} \times p_\lambda(y(t) \mid s(t)=j)\beta_t(j) \quad (7)$$

$$p_\lambda(s(t)=j, m(t)=k \mid y(t)) = \sum_{i=1}^{L} \alpha_{t-1}(i)\, a_{ij} c_{k|j} \times p_\lambda(y(t) \mid s(t)=j, m(t)=k)\beta_t(j) \quad (8)$$

여기서

$\alpha_t(j) = \sum_{i=1}^{L} \alpha_{t-1}(i)\, a_{ij}\, p_\lambda(y(t) \mid s(t)=j)$,

$\beta_t(j) = \sum_{i=1}^{L} \beta_{t-1}(i)\, a_{ij}\, p_\lambda(y(t+1) \mid s(t+1)=i)$이고,

$p_\lambda(y(t) \mid s(t)=j) = \sum_{k=1}^{M} c_{k|j}\, p_\lambda(y(t) \mid s(t)=j, m(t)=k)$

이다. 단 $a_0(i)=\pi_i$이고 모든 j에 대해 $\beta_T(j)=1$이다.

유색잡음은 마코프 모델에 따라서 변하는 파라미터를 갖는 AR 모델인 HFM에 의해서 다음과 같이 모델링된다.

$$v(t) = C^T_{h(t)}\, v(t-1) + w_{h(t)}(t) \quad (9)$$

여기서 $h(t) \in \{1,2,\cdots,N\}$은 상태를 나타내고, $v(t-1) = [v(t-1)\cdots v(t-q)]^T$는 q개의 관측데이터로 구성된 벡터이며, $w_{h(t)}(t)$는 평균이 0이고 분산이 $\sigma^2_{w,h(t)}$인 가우시안 프로세스이다. 잡음에 대한 HFM의 파라미터 $\lambda_v = \{\tilde{a}, c, \sigma^2_w\}$는 위의 방법과 같이 학습음성신호로부터 Baum-Welch 알고리즘을 사용하여 추정할 수 있다. 여기서 $\tilde{a} = \{\tilde{a}_{ij}\}$, $C = \{C_j\}$, $\sigma^2_w = \{\sigma^2_{w,j}\}$이고 $i, j = 1, \ldots, N$이다.

이것을 기반으로 잡음에 오염된 음성신호는 어떻게 표현되는지 살펴보자.

$$z(t) = y(t) + v(t) \quad (10)$$

음성에 대한 AR 모델과 잡음에 대한 AR 모델을 결합하면 다음과 같은 복합 마코프상태 $\theta(t) = \{m(t), s(t), h(t)\}$에 따른 상태방정식으로 표현할 수 있게 된다.

$$x(t) = F(\theta(t))\, x(t-1) + G\, e(\theta(t)) \quad (11)$$

$$z(t) = H^T x(t) \quad (12)$$

여기서 $x(t) = \begin{bmatrix} y(t) \\ v(t) \end{bmatrix}$, $e(\theta(t)) = \begin{bmatrix} e_{m(t)|s(t)}(t) \\ w_{h(t)}(t) \end{bmatrix}$이고 $F(\theta(t)) = \begin{bmatrix} \Phi(m(t), s(t)) & 0 \\ 0 & F_v(h(t)) \end{bmatrix}$, $G = \begin{bmatrix} g_y & 0 \\ 0 & g_v \end{bmatrix}$, $H = \begin{bmatrix} H_y \\ H_v \end{bmatrix}$이다. 한편 음성신호의 인가신호 $e_{m(t)|s(t)}(t)$와 잡음의 인가신호 $w_{h(t)}(t)$는 서로 비상관되어 있으므로 다음과 같이 표현할 수 있다.

$$Q(\theta(t)) \equiv E\{\, e(\theta(t))\, e^T(\theta(t))\} = \begin{bmatrix} \sigma^2_{m(t)|s(t)} & 0 \\ 0 & \sigma^2_{w,h(t)} \end{bmatrix} \quad (13)$$

Ⅲ. 음성신호의 추정

식 (12), (13)과 같이 상태방정식이 주어졌을때, $x(t)$의 값을 추정하게 되면 그중 $y(t)$가 향상된 음성신호가 된다. $x(t)$를 구하기 위해 혼합 상호복합모델 (MIMM) 알고리즘은 식 (12), (13)으로 주어지는 각각의 $L \times M \times N$개의 다른 모델에 대한 $L \times M \times N$개의 칼만 필터를 사용하여 구성된다. 매 단계의 시작에 각 필터의 초기 값들이 따로 주어지고, 매 단계의 끝에는 각 필터들에 의해 출력된 추정 값들을 결합하여 최종적인 추정 값을 생성하게 된다. 매 단계의 입력으로는 $L \times M \times N$개의 모델에 해당하는 조건부 추정 값,

$$\hat{x}_{\alpha\beta\gamma}(t-1) \equiv E\{x(t-1) \mid \theta_{\alpha\beta\gamma}(t-1),\ z(t-1)\},$$
$$1 \le \alpha \le M,\ 1 \le \beta \le L,\ 1 \le \gamma \le N. \quad (14)$$

과 그것에 해당하는 공분산, $P_{\alpha,\beta,\gamma}(t-1)$ 그리고 다음 식으로 주어지는 $L \times M \times N$개의 가중치, $p(Hist(t,o) \mid z(t))$이 주어진다. 여기서 $Hist(t,o) = \{\theta_{o_0}(0), \theta_{o_1}(1), \cdots \theta_{o_t}(t)\}$, $o = (o_0, o_1, \cdots o_t)$는 모든 가능한 복합 상태로부터 나올 수 있는 임의의 관찰벡터 열을 나타내며, θ_{o_t}는 시간 t에서 $o_t = i, j, k$인 복합상태, $\{m(t)=i, s(t)=j, h(t)=k\}$를 의미한다. 해당하는 확률밀도함수는 다음과 같은 정규분포를 갖는다고 가정한다.

$$p(x(t-1) \mid \theta_{\alpha\beta\gamma}(t-1),\ z(t-1)) \approx N(\hat{x}_{\alpha\beta\gamma}(t-1),\ P_{\alpha\beta\gamma}(t-1)) \quad (15)$$

MIMM 알고리즘의 작동은 다음의 4과정으로 이루어진다.

(1) 입력 생성 과정 (혼합 과정)

각 모델 $\{\theta_{ijk}(t),\ 1 \le i \le M, 1 \le j \le L, 1 \le k \le N\}$에 대응하는 칼만 필터의 초기 조건들은 이전 단계에서의 각

필터들에 의한 추정 값들을 혼합함으로써 구할 수 있다.

$$\boldsymbol{x}^0_{ijk}(t-1) = E\{\boldsymbol{x}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\}$$
$$= \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} E\{\boldsymbol{x}(t-1)|\theta_{ijk}(t),\theta_{\alpha\beta\gamma}(t-1),$$
$$\boldsymbol{z}(t-1)\}\times P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\} \quad (16)$$

여기서 $P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\}$는 혼합확률이고, $\theta_{\alpha\beta\gamma}(t-1)$이 주어진다면 $\boldsymbol{x}(t-1)$과 $\theta_{ijk}(t)$는 서로 독립이므로 다음의 식을 얻을 수 있게 된다.

$$\boldsymbol{x}^0_{ijk}(t-1) = E\{\boldsymbol{x}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\}$$
$$= \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} E\{\boldsymbol{x}(t-1)|\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}$$
$$\times P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\} \quad (17)$$

그러면 식 (14)로부터 다음의 식을 얻을 수 있다.

$$\boldsymbol{x}^0_{ijk}(t-1) = \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} \hat{\boldsymbol{x}}_{\alpha\beta\gamma}(t-1)$$
$$\times P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\} \quad (18)$$

여기서 우측항의 가중치 확률 $P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\}$은 Bayes rule을 사용하여 다음의 식으로 구한다.

$$P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\}$$
$$= \frac{P\{\theta_{ijk}(t)|\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}P\{\theta_{\alpha\beta\gamma}(t-1)|\boldsymbol{z}(t-1)\}}{P\{\theta_{ijk}(t)|\boldsymbol{z}(t-1)\}} \quad (19)$$

복합상태 $\theta_{ijk}(t)$에서 잡음에 오염되지 않은 음성신호의 상태, $\{m(t)=i,\ s(t)=j\}$와 잡음의 상태, $\{h(t)=k\}$가 서로 독립이므로 $P\{\theta_{ijk}(t)|\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}$을 다음과 같이 표현할 수 있다.

$$P\{\theta_{ijk}(t)|\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}$$
$$= P\{m_i(t),\ s_j(t)|\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}$$
$$\times P\{h_k(t)|\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}$$
$$= P\{s_j(t)|s_\beta(t-1),\ \boldsymbol{z}(t-1)\}$$
$$\times P\{m_i(t)|s_j(t),\ \boldsymbol{z}(t-1)\}$$
$$\times P\{h_k(t)|h_\gamma(t-1),\ \boldsymbol{z}(t-1)\} = a_{\beta j}\,c_{ij}\,\tilde{a}_{\gamma k} \quad (20)$$

위의 결과로부터 식 (19)를 나타내 보면 다음과 같다.

$$P\{\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\}$$
$$= \frac{a_{\beta j}\,c_{ij}\,\tilde{a}_{\gamma k}P\{\theta_{\alpha\beta\gamma}(t-1),\ \boldsymbol{z}(t-1)\}}{\sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} a_{\beta j}\,c_{ij}\,\tilde{a}_{\gamma k}P\{\theta_{\alpha\beta\gamma}(t-1)|\boldsymbol{z}(t-1)\}} \quad (21)$$

각각에 해당하는 추정값의 공분산들은 비슷한 방법으로 계산될 수 있으며 다음과 같이 나타낼 수 있다.

$$\boldsymbol{P}^0_{ijk}(t-1) = \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} P(\theta_{\alpha\beta\gamma}(t-1)|\theta_{ijk}(t),\ \boldsymbol{z}(t-1))$$
$$\times[\boldsymbol{P}_{\alpha\beta\gamma}(t-1) + (\hat{\boldsymbol{x}}_{\alpha\beta\gamma}(t-1) - (\boldsymbol{x}^0_{\alpha\beta\gamma}(t-1))$$
$$\times(\hat{\boldsymbol{x}}_{\alpha\beta\gamma}(t-1) - (\boldsymbol{x}^0_{\alpha\beta\gamma}(t-1)))] \quad (22)$$

(2) 칼만 필터링 과정

각 칼만 필터의 입력으로서 $(\boldsymbol{x}^0_{ijk}(t-1),\ \boldsymbol{P}^0_{ijk}(t-1))$이 주어지고 출력 $(\hat{\boldsymbol{x}}(t),\ \boldsymbol{P}_{ijk}(t))$는 칼만 필터에서 계산하는데 그 과정은 식 (11), (12)로부터 다음의 식으로 표현할 수 있다.

$$\boldsymbol{M}_{ijk}(t)$$
$$= F(\theta_{ijk}(t))\ \ \boldsymbol{P}^0_{ijk}(t-1)\,F^T(\theta_{ijk}(t)) + GQ(\theta_{ijk}(t))\,G^T \quad (23)$$

$$\boldsymbol{K}_{ijk}(t) = \boldsymbol{M}_{ijk}(t)H[H^T\ \boldsymbol{M}_{ijk}(t)]^{-1} \quad (24)$$

$$\hat{\boldsymbol{x}}_{ijk}(t) = F(\theta_{ijk}(t))\ \ \boldsymbol{x}^0_{ijk}(t-1)$$
$$+\ \boldsymbol{K}_{ijk}(t)[z(t) - H^T F(\theta_{ijk}(t))\ \boldsymbol{x}^0_{ijk}(t-1)] \quad (25)$$

$$\boldsymbol{P}_{ijk}(t) = \boldsymbol{M}_{ijk}(t) - \boldsymbol{K}_{ijk}(t)\ H^T\ \boldsymbol{M}_{ijk}(t) \quad (26)$$

여기서

$$\boldsymbol{P}^0_{ijk}(t-1) = E\{(\boldsymbol{x}(t-1) - \boldsymbol{x}^0_{ijk}(t-1))$$
$$(\boldsymbol{x}(t-1) - \boldsymbol{x}^0_{ijk}(t-1))^T|\theta_{ijk}(t),\boldsymbol{z}(t-1)\} \quad (27)$$

$$\boldsymbol{x}^0_{ijk}(t-1) = E\{\boldsymbol{x}(t-1)\mid\theta_{ijk}(t),\ \boldsymbol{z}(t-1)\} \quad (28)$$

이다. 위의 과정에서 다음 식과 같은 가우시안 분포 가정이 사용되었다.

$$p(\boldsymbol{x}(t-1)|\theta_{ijk}(t-1),\ \boldsymbol{z}(t-1)) \approx$$
$$N(\boldsymbol{x}^0_{ijk}(t-1),\ \boldsymbol{P}^0_{ijk}(t-1)) \quad (29)$$

(3) 가중치 확률 계산 과정

가중치 확률은 최종 출력을 구하기 위해 필요한데 그 계산식은 다음과 같다.

$$P\{\theta_{ijk}(t)|\ \boldsymbol{z}(t)\}$$

$$= \frac{P\{z(t) \mid \theta_{ijk}(t),\ z(t-1)\}P\{\theta_{ijk}(t) \mid z(t-1)\}}{P\{z(t) \mid z(t-1)\}}$$

$$= \frac{P\{z(t) \mid \theta_{ijk}(t),\ z(t-1)\}}{P\{z(t) \mid z(t-1)\}}$$

$$\times \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} a_{\alpha\beta}\, c_{i|j}\ \tilde{a}_{\gamma k} P\{\theta_{\alpha\beta\gamma}(t-1) \mid z(t-1)\} \quad (30)$$

여기서 $P\{z(t) \mid \theta_{ijk}(t),\ z(t-1)\}$

$$= \frac{1}{\sqrt{2\pi \sum_{ijk}}} \exp\left(-\frac{1}{2}(\tilde{z}(t))^{T} \sum_{ijk}^{-1}(\tilde{z}(t))\right) = N_{ijk}$$

$\tilde{z}(t) = z(t) - \boldsymbol{H}^{T}\boldsymbol{F}(\theta_{ijk}(t)\ \ \boldsymbol{x}^{0}{}_{ijk}(t-1)$은 $\theta_{ijk}(t)$번째 칼만 필터의 이노베이션 (innovation) 시퀀스이다. 따라서 가중치는 다음 식과 같이 이전 가중치를 사용해 반복적으로 구할 수 있다.

$$P\{\theta_{ijk}(t) \mid z(t)\}$$

$$= D_t N_{ijk} c_{i|j} \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} a_{\alpha\beta}\ \tilde{a}_{\gamma k} P\{\theta_{\alpha\beta\gamma}(t-1) \mid z(t-1)\} \quad (31)$$

여기서 D_t는 시간 t에서의 가중치이고, 시작확률값은 적절한 값으로 초기화된다.

(4) 출력 생성 과정

최종 추정치와 그에 해당하는 공분산은 다음의 식과 같이 각 칼만 필터들의 출력의 가중 합으로 주어진다.

$$\tilde{\boldsymbol{x}}(t) = \sum_{i=1}^{M}\sum_{j=1}^{L}\sum_{k=1}^{N} \hat{\boldsymbol{x}}_{ijk}(t)\, P\{\theta_{ijk}(t) \mid z(t)\} \quad (32)$$

$$\boldsymbol{P}(t) = \sum_{i=1}^{M}\sum_{j=1}^{L}\sum_{k=1}^{N} P(\theta_{ijk}(t) \mid z(t))[\ \boldsymbol{P}_{ijk}(t) + (\ \hat{\boldsymbol{x}}_{ijk}(t) - \hat{\boldsymbol{x}}(t))\ (\ \hat{\boldsymbol{x}}_{ijk}(t) - \hat{\boldsymbol{x}}(t))^{T}] \quad (33)$$

MIMM 알고리즘은 위의 4 과정을 매 단계마다 반복적으로 수행하여 향상된 음성신호를 얻게 된다. 하지만 식 (31)의 가중치 확률계산에 많은 계산량을 필요로 하므로 실제 구현에 있어서 다음과 같은 효율적인 알고리즘을 사용하여 계산하였다. 가중치 확률계산에서 필요한 N_{ijk}, $P\{\theta_{\alpha\beta\gamma}(t-1) \mid z(t-1)\}$값이 아주 작을 때는 확률값을 무시할 수 있기 때문에 이 값이 어떤 문턱값을 넘는 경우에만 가중치 확률을 계산함으로써 계산량을 상당히 줄일 수 있었다. 따라서 식 (31)은 다음과 같이 표현될 수 있다.

$$p(\theta_{i^{*}j^{*}k^{*}}(t) \mid z_1^{t}) =$$

$$D_t \cdot N_{i^{*}j^{*}k^{*}} c_{i^{*}|j^{*}} \sum_{\alpha^{*}}\sum_{\beta^{*}}\sum_{\gamma^{*}} a_{\beta^{*}j^{*}} \tilde{a}_{\gamma^{*}k^{*}} p(\theta_{\alpha^{*}\beta^{*}\gamma^{*}}(t-1) \mid z(t-1))$$

#, *는 다음과 같이 선택된 상태수를 나타낸다.

$$i^{*}j^{*}k^{*} \ <= \ \text{if} \ \ N_{ijk} \geq threshold1$$

$$\alpha^{*}\beta^{*}\gamma^{*} \ <= \ \text{if} \ \ p\{\theta_{\alpha\beta\gamma}(t-1) \mid z(t-1)\} \geq threshold2 \quad (34)$$

$$\hat{x}(t) = \sum_{i}\sum_{j}\sum_{k} \hat{\mathrm{x}}_{i\,j\,k}(t) p(\theta_{i^{*}j^{*}k^{*}}(t) \mid z(t))$$

최종적으로 추정된 음성신호는 다음의 식과 같이 p-1 샘플만큼 지연시킨 $\hat{x}(t)$의 p번째 요소가 된다.

$$\hat{y}(t) = \left| \underbrace{0\ldots010\ldots0}_{p-1,\ p,\ q} \right| \hat{x}(t+p-1)$$

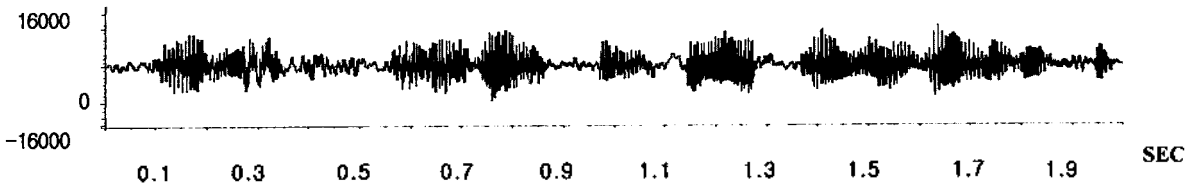
IV. 실험결과

여러 경우의 입력 신호대잡음비 (5 dB, 10 dB, 15 dB)에 대해서 신호향상 실험을 행하였다. 잡음은 유색잡음의 일종인 배블 (babble)잡음과 자동차 소음 등으로 하였다. 잡음의 경우 상태모델의 개수는 4, AR모델의 차수는 15차인 HFM로 모델링하였다. 잡음에 오염되지 않은 음성신호의 경우 HFM에 대한 파라미터를 추정하기 위하여 학습은 10명의 남성화자와 5명의 여성화자에 의해 발음된 50개의 문장을 이용하여 수행하였다. 음성신호 향상 실험은 학습신호에 사용된 화자가 아닌 화자에 의해 발음된 문장에 대해서 이루어졌다. 음성신호를 8kHz로 표본화했고 AR 모델의 차수는 10을 사용했으며 상태모델의 개수는 5, 모델별 혼합 (mixture)은 5개로 하였다.

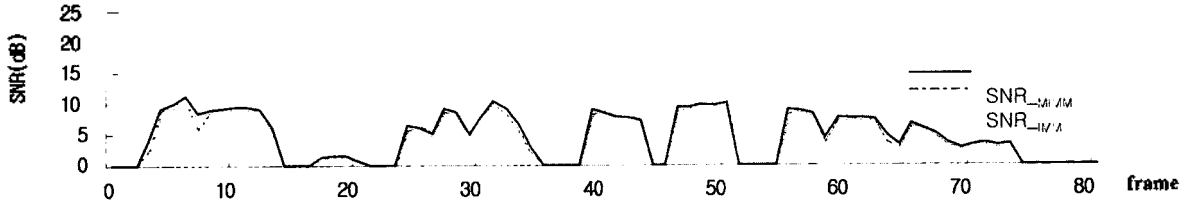
여기서 식 (34)의 문턱값 1, 2는 실험적으로 결정하였으며, 대략 0.3-0.8 정도일 때 하나 또는 두 개의 상태가 선택된다. 따라서 문턱값이 음성신호 향상에 별 영향이 없음을 알 수 있다. 학습에 사용된 신호들은 비슷한 정도의 크기를 갖는 전력 정규화 (power normalization)를 행하였다. 신호향상 실험에 사용된 신호들도 전력 정규화를 사용하여 학습데이터들과 비슷한 크기를 갖도록 하였다.

표 1. 구간 신호대잡음비 성능
Table 1. Segmental SNR performance.

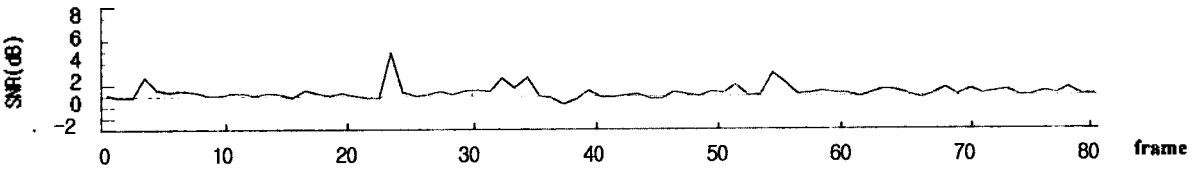
SNR	배블잡음		자동차소음	
	제안한방법	기존방법	제안한방법	기존방법
5	9.87	8.45	9.98	8.78
10	14.93	13.82	15.12	14.23
15	18.78	17.54	18.98	17.99



(a) 오염된 음성신호 (5dB 자동차소음)
(a) Contaminated speech data (SNR=10dB)

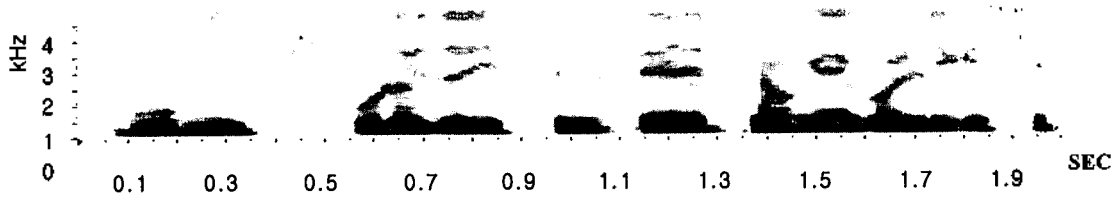


(b) 프레임에 따른 신호대잡음비
(b) Frame-based SNR

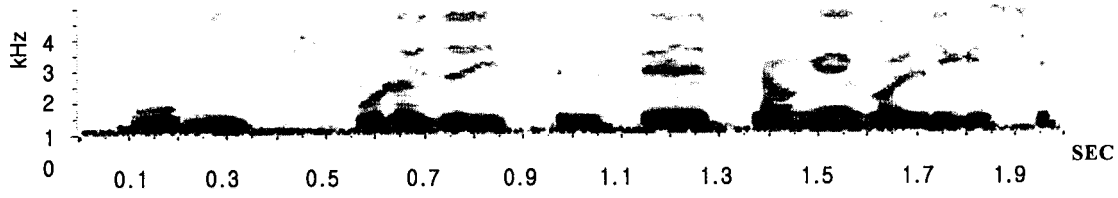


(c) 제안된 방법의 신호대잡음비 성능향상 ($SNR_{MIMM} - SNR_{IMM}$)
(c) SNR improvement ($SNR_{MIMM} - SNR_{IMM}$)

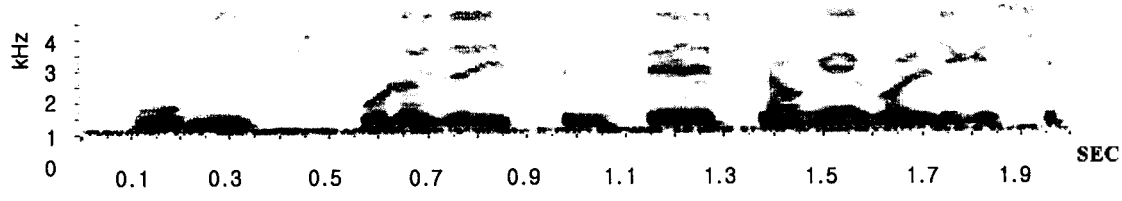
그림 1
Fig. 1



(a) 오염되지 않은 음성신호
(a) Original speech



(b) 오염된 음성신호
(b) Noisy speech



(c) 제안한기법
(c) MIMM based method

그림 2. 향상된 신호의 스펙트로그램 (5dB 자동차소음)
Fig. 2. Spectrogram of enhanced signal (5dB car noise).

기억장소와 계산량을 줄이기 위해 Baum-Welch 알고리즘 대신 세그멘탈 k-평균 알고리즘[6]을 사용하였다.

표 1에서는 배블잡음과 자동차 소음에 오염된 음성신호의 성능향상을 보여주기 위해 기존방법인 IMM기법과 본 논문에서 제안한 기법에 대한 구간 신호대잡음비에 대한 결과가 제시되어 있다. 구간 신호대잡음비는 프레임 크기를 200 샘플로 하여 프레임마다의 신호대잡음비를 계산한 후 이들을 합하고 프레임 개수로 나누어 평균을 구하였다. 묵음구간에 해당하는 프레임들을 배제하기 위해서 프레임의 신호대잡음비가 −15 dB 이하인 프레임은 계산에서 제외하였다. 제안한 기법이 IMM기반 기법보다 구간 신호대잡음비에서 약 0.3 dB-1.2 dB 가량 개선된 성능을 보임을 알 수 있다.

그림 1, 2에서는 자동차 잡음에 오염된 음성신호에 대해서 제안한 기법에 대한 향상된 음성신호에 대한 스펙트로그램을 보였다. 프레임 신호대잡음비는 프레임 크기 200 샘플을 사용했고 스펙트로그램은 크기가 256 샘플인 Hanning 창을 사용하였다. 전반적으로 안정된 신호구간에서는 기존 알고리즘인 IMM에 비해 차이가 크지 않으나 전이 구간에서는 상당한 신호대잡음비 차이를 보임을 알 수 있다. 이것은 알고리즘의 적응 성능의 차이와 모델링의 정확성의 차이에 기인하는 것으로, 제안한 기법이 기존기법에 비해 좋은 성능을 보임을 확인할 수 있었다.

V. 결 론

본 논문에서는 유색잡음 중에서 배블 잡음과 자동차 소음에 오염된 음성신호의 향상을 위해 혼합 IMM에 기반한 효율적인 순환 알고리즘을 제안하였다. 오염되지 않은 음성신호는 혼합 HFM, 잡음은 하나의 HFM로 모델링했으며, 모델의 파라미터는 학습 음성신호로부터 추정하였다. 제안한 방법에서 음성신호의 추정값은 서로 상호작용하는 병렬의 칼만필터들의 가중합으로 주어지며, 가중치는 오염된 음성신호가 주어졌을 때 복합상태의 사후확률값이다. 제안한 방법을 통해 모델링을 더 정확하게 함으로써 계산량의 큰 증가없이 기존의 방법보다 개선된 음성향상 성능을 얻을 수 있었다.

참 고 문 헌

1. R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, **28**, 137-145, Apr. 1980.
2. J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Acoust., Speech, Signal Process.*, **39**, 1732-1742, August 1991.
3. Y. Ephraim, "Statistical model based speech enhancement system," *Proc. IEEE*, **80**, 1526-1555, 1992.
4. Y. Ephraim, D. Malah, and B. Juang, "On the application of hidden Markov models for enhancing noisy speech," *IEEE Trans. Acoust., Speech, Signal Process.*, **37**, 1846-1856, Dec. 1989.
5. Y. Ephraim, "A Bayesian estimation approach for speech enhancement using hidden Markov models," *IEEE Trans. Acoust., Speech, Signal Process.*, **40**, 725-735, April 1992.
6. H. Sheikhzadeh and L. Deng, "Waveform-based speech recognition using hidden filter models: parameter selection and sensitivity to power normalization," *IEEE Trans. Acoust., Speech, Signal Process.*, **2**, 80-89, Jan. 1994.
7. K. Y. Lee and K. Shirai, "Efficient recursive estimation for speech enhancement in colored noise," *IEEE Signal Processing Letters*, **3**, 196-199, July 1996.
8. J. B. Kim, K. Y. Lee and C. W. Lee, "On the application of the interacting multiple model algorithm for enhancing noisy speech," *IEEE Trans. Audio and Speech Processing*, **8**, 349-352, May 2000.

저자 약력

● 강 상 기 (SangKi Kang)

1992년 2월: 창원대학교 전자공학과 졸업 (공학사)
1997년 2월: 서울대학교 대학원 전자공학과 졸업 (공학석사)
1997년~2002년: 서울대학교 대학원 전기·컴퓨터공학부 졸업 (공학박사)
※ 주관심분야: 음성 신호처리, 오디오/음성 코딩, 적응 신호처리

● 백 성 준 (SeongJoon Baek)
한국음향학회지 제16권 제2호 참조
현재: 전남대학교 전자컴퓨터정보통신공학부 조교수

● 이 기 용 (Ki Yong Lee)
한국음향학회지 제15권 제3E호 참조
현재: 숭실대학교 정보통신과 교수

● 성 굉 모 (Koeng-Mo Sung)
한국음향학회지 제14권 제2호 참조
현재: 서울대학교 전기·컴퓨터공학부 교수